Town James S (Orcid ID: 0000-0003-4811-0076)

Automatic Peak Assignment and Visualisation of Copolymer Mass Spectrometry Data Using the "Genetic Algorithm"

James S Town, Yuqui Gao, Ellis Hancox, Evelina Liarou, Ataulla Shegiwal, Christophe J Atkins, David Haddleton**

Introduction

Copolymers represent a wide range of materials encompassing different chemical, mechanical and thermal properties. Properties can be somewhat tuned by altering the structure and compositional makeup of the copolymer chain[1-6]. It is due to the relationship between synthetic methodology and tuneable properties that copolymers have found such diverse and essential applications[7-11]. Copolymers of vinyl monomers maybe synthesised by a wide variety of polymerisation methods including; catalytic chain transfer polymerisation (CCTP)[12-14], atom transfer radical polymerisation (ATRP)[15-17], ionic polymerisation[18,19], reversible addition-transfer chain-transfer polymerisation (RAFT)[20-22], and sulphur free RAFT[23-26]. These polymerisation methods can lead to different challenges in mass spectrometry, from examination of labile end groups[27-30] (such as in ATRP and RAFT) to higher dispersities leading to a wide *m/z* range to cover[31-34] (such as in CCTP). These challenges become even more complex, often to the point of becoming intractable and unsolvable, in the case of copolymers due to the enormous number of different molecular species present in a material. Hence, improvements of how we approach copolymer analysis is of on-going interest.

Polymers, are mixtures of different molecular species, which become increasingly more complex when two or more monomers are introduced to the system. This can make characterisation extremely difficult, especially when it comes to exact determination of what species exist and in what relative quantity. Previous approaches have included 2 dimensional NMR methodology[35-37], LC/MS[38,39], and different 2 dimensional chromatography[40-42]. All of these approaches are complex and may not always give the detailed understanding of these copolymer materials that is required.

Data processing techniques for mass spectrometry data has become more of a relevant field as the complexity of data has increased, with more powerful mass spectrometers being utilised, and the technique has gained wider usage in industrial fields. Polymeromics is no exception to this; however, as with many aspects of mass spectrometry, this sub-field lags behind the neighbouring fields (proteomics, petroleomics, lipidomics etc.).

Kendrick Mass Defect plots have proven invaluable to other areas of mass spectrometry, such as proteomics[43,44], and have been applied rigorously to polymers. This includes such improvements as fractional base units[45,46] and slicing[47], which have, or are likely to in the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rcm.8654

near future, greatly improved their application to copolymer analysis[48]. The benefit of KMD plots is that they are applied to all the peaks in the sample, displaying all structures which have the base unit as horizontal lines. The downside, however, is that while it does simplify assignment by processing peaks into lines with the same KMD, the assignment still has to be carried out manually, or by separate automation.

Mass Remainder Analysis (MaRA) as proposed by Nagy et al[49], is similar to Kendrick Mass Defect; however. it is much more simplified, utilizing the division of a peak by one of the repeat units to allow for the separation of species into visual horizontal rows. Hence, whilst it does give visualisation of the copolymer, it still requires manual assignment, and so this technique is very similar to Kendrick mass defect analysis[50] and hence shares similar drawbacks.

Willemse et al have utilizing MALDI data to develop contour plots for copolymer fingerprinting, using an array-based assignment. They were able to track the progress of a polymerisation of a block copolymer and demonstrate the contour plot changing through their spectra[51]. They, however, ran into issues with multiple assignments for a single peak, an issue that arises from the lower resolution mass spectrometers available at the time of the study and the array methodology being used due to the lower computational power available at the time.

The genetic algorithm has been applied to mass spectrometry analysis of metabolite systems before. This methodology is used for the purpose of predicting markers involved in diagnosis of cancer from patients, alongside more traditional principal component analysis[52]. This is a very different purpose from the direct peak assignment that we will report in this work.

Genetic algorithm analysis has also been applied to tandem mass spectrometry data of glycosaminoglycans. This analysis allowed for high throughput structural determination of these species by reducing the R groups to a binary sequence. This analysis provides fast and accurate analysis for these structures; however, such a binary sequence may be more difficult to implement on synthetic polymer samples due to the range of monomers[53].

In this current work, we describe our use of the genetic algorithm to automatically assign peaks in MALDI-ToF data for copolymer samples. As an example of the usefulness of the generated output, the data is used to generate simple visualisations of complex copolymer spectra, which will allow non-expert users to analyse copolymer samples. We believe that the genetic algorithm peak assignment could lead to more advanced, automated copolymer analysis in the future.

Methods

MALDI-ToF

MALDI-ToF experiments were carried out using a Bruker (Bremen, Germany) Autoflex mass spectrometer, equipped with a 337 nm N_2 laser, operating at 21 kV acceleration voltage in reflectron positive mode. Samples were prepared in THF at (10 mg mL^{-1}) with sodium iodide salt (1 mg mL^{-1}) and a DCTB matrix (40 mg mL^{-1}), the only exception being the Styrene-

Methyl Methacrylate copolymer which was prepared with silver trifluoroacetate (1 mg mL^{-1}) as well as the sodium iodide salt. The solution was then spotted onto a MTP 382 ground steel target plate for analysis.

Mathematics and Scripting

Matlab was utilized to script all the data analysis, including the production of the graphs shown throughout this article. In order to generate automated peak picking we utilized the genetic algorithm function found in the global optimisation toolbox. This algorithm was used due to it allowing for integer constraints; the parameters to allow for the fastest, correct, assignment are described in the supporting information. Equation 2, shown below, has the mass values of end groups (E), monomer 1 (M_1), monomer 2 (M_2) and ionising salt (S) which are all known given a single manually assigned peak. The genetic algorithm therefore is utilised to find the minimum value of error by adjusting the number of monomer 1 and monomer 2 units (N_1 and N_2):

$$Error = Theoretical\ Mass - Experimental\ Mass \quad (1)$$

$$Error = E + N_1 \times M_1 + N_2 \times M_2 + S - Experimental\ Mass \quad (2)$$

In a perfectly calibrated mass spectrum we would be able to minimise this equation to 0. However, no mass spectrometry is ever perfect, and therefore there will always be an associated error. The script, therefore, includes an adjustable error cut-off which, after it has finished assigning all peak,s is then used to remove any assignments not satisfying this error. We recommend an error cut-off of 0.1 *m/z* units or below, as this is perfectly achievable with even external calibration in relatively low resolution ToF instruments.

Once a peak has been assigned, to allow for better representation of the intensity in the mass spectra, the script attempts to find all the isotopic peaks which relate to the assigned peak and sum their intensities. This is to avoid higher molecular weight peaks being under represented by the intensity of their monoisotopic mass, which is used to calculate the assignment, as this is not the highest intensity peak for carbon-based polymers with molecular masses above ~2000 Daltons depending on the chemical formula. This is achieved by attempting to find a peak, which is both 1 *m/z* unit higher, within assignment error, and has an intensity which is less than a selected multiple of that of the original peak. This intensity factor is not set to allow for adjustment for samples with halides, or other elements with more complex isotopic distributions. Peaks which are determined to be isotopes of a previous peak are discounted from being assigned later, and are hence not put through the genetic algorithm. The genetic algorithm is by far the most computationally expensive part of the code; therefore, discounting these peaks before assignment allows for less processing time.

Results

Optimization of Genetic Algorithm Parameters

The parameters used in the genetic algorithm were optimized using a poly (methyl acrylate – ethyl acrylate) statistical copolymer, with a 50/50 ratio between the two monomers, synthesized by Cu (0) mediated SET-LRP. The optimizations carried out are presented in the supporting information, including the specifications for the laptop used to carry out the Matlab script, and the final parameters utilized for the algorithm, which are in the table below:

Initial Population	Elite Children	Function Tolerance	Max Generations
Permeations/4	0.7 * Previous Population	10^{100}	40

Permeations is a value calculated as the number of all possible combinations of the two monomers calculated as below:

$$DP_{pred} = \frac{\frac{m}{z} maximum}{M_{maximum}} \quad (3)$$

$$Permeations = \frac{DP_{pred}!}{\big(DPpred - N(monomers)\big)!} \quad (4)$$

where DP_{pred} is representing the predicted maximum degree of polymerization that the chains can take, calculated by dividing the maximum *m/z* value in the dataset by the mass of the highest mass monomer being used for assignment. Permeations is therefore calculated using the predicted maximum degree of polymerisation (DP_{pred}) and the number of different monomers (N (monomers)). By making the initial population alter based on the number of possible results, the algorithm does not have to be manually altered for polymers with more or less possibilities, giving accurate results without user input.

With the current optimization of this algorithm, the Matlab script currently takes 17 seconds on a >900 peak dataset, reducing it to 110 species with good repeatability.

MA/EA Statistical Copolymers

The automatic peak assignment by the genetic algorithm is used to generate a heat map with N_1 on the x axis, N_2 on the y axis, and a transformed intensity at the colour gradient. This provides visualisation of the copolymers, allowing for simple qualitative comparison. The heat maps for methyl acrylate-co-ethyl acrylate with monomer ratios of 50/50, 60/40, 70/30, 80/20 and 90/10 are distinct in their overall shape as the more MA in the monomer ratio compared with EA the heat map appears to have a shallower gradient. The heat maps also provide a visual diagnostic for the assignment as the distribution provided on the heat map has no gaps, which could imply peaks missed by the algorithm, or higher intensity points lying outside the main distribution, which would imply peaks which were assigned

incorrectly. The 70-30 copolymer displays this, as its width is due to a miss-assigned peak on the very far right of the heat map (Figure 3).

It is therefore possible to use the heat map to find the peaks which have been missed or which have been assigned incorrectly in the original spectra. This allows for the visualisation as a diagnostic tool for the genetic algorithm assignment. The assignment is reliant on good calibration, as this will minimise the error that is set as a cut-off for correct assignments. The example MA/EA 60/40 heat map, in Figure 4, shows this effect of poor calibration. In this case it would appear that several isotopic peaks have been assigned as real species. This indicates that the calibration led to them not being correctly assigned as isotopes, and hence they were not removed from the potential assignments. Falsely assigned isotopic peaks also have the downside of causing the relative intensities in the heat map, and the absolute intensities in the genetic algorithm output, to be less representative of the real data, as the isotopic distribution is not correctly summed into the real assigned peak.

Comparing methyl acrylate-co-ethyl acrylate copolymers made by two different synthetic chemists using two different and distinct forms of copper mediated living radical polymerization (one photo mediated[54-56], the other using a copper(0) wire system[57-60]), we can draw some simple conclusions about the synthesis in a qualitative manner. By examining the heat maps for a copolymer with a 50/50 mole% composition side by side we can see that the copper wire system was more controlled, in that the distribution of the copolymer spectra seems to be less dispersed. This form of examination is a new way of looking at synthetic copolymerizations, as we can examine an under-evaluated area of synthetic control, the control over the composition of the chains.

Analysis of MMA/EMA diblock copolymers

When a 10 MMA 10 EMA diblock, synthesised by via a combination of CCTP[12,13] and sulphur-free raft (SF RAFT)[23-25], was analysed by MALDI-ToF its spectrum had interesting features as it did not contain a normal compositional distribution with narrow dispersity. The spectrum was then analysed with the genetic algorithm and displayed as a heat map. One of the issues in this spectrum is how broad it is; it is found that in spectra over this range of masses it is difficult to get a very high accuracy of calibration. Hence the assignment error is higher in some of the real peaks, which, when accounted for, leads to some miss-assignments. Lower abundance species have overlapping isotopic distributions higher abundance species, and hence some assignments are also lost when the intensity cut-off factor of our isotopic distribution assignment is too high. This is because peaks which come after the lower abundance species can be assigned as isotopes of those lower abundance peaks, similar to the MA/EA system.

The heat map in Figure 7 shows that the sample contains high amounts of PMMA homopolymer. This implies that the incorporation of the macromonomer into a block copolymer was incomplete, even though the monomer conversion was taken to a high percentage (>95%). The polymer has a broad dispersity, around 1.7, which could mean that higher molecular weight chains contain more of the EMA than the MMA polymers; however, the limitations of the mass spectrometer prevents the accurate analysis of copolymer distributions with >10,000 molecular weight. The other significant difference between this and the previous example is the greatly increased number of molecular species. This is because the number of copolymer species observed in a diblock copolymer sample is related to the molecular weight distribution of the second block of the diblock, which is likely also to be very broad.

This displays the importance of mass spectrometry relative to bulk measurements which are traditionally used in polymer characterisation, such as 1 dimensional NMR, which would not be able to show this homopolymer problem, instead providing an average monomer incorporation in all polymeric chains. Using MALDI-ToF-MS, in collaboration with the genetic algorithm peak assignment, we are able to display the data with ease.

Analysis of a MMA-Styrene statistical copolymer

A methyl methacrylate – styrene statistical copolymer synthesised by free radical in bulk (supporting information) demonstrates some of the effect of reactivity ratios on the number of observed species in the mass spectrum (Figure 8). The reactivity ratios of MMA – Styrene copolymers have been shown to be r_{MMA} = 0.51 and $r_{Styrene}$ = 0.49;[61] this implies that the reaction tends towards a slightly alternating sequence. This, therefore, would lead to a reduction in the number of species, as alternating polymers would have a maximum of 3 species per degree of polymerization. We can observe, using the genetic algorithm assignment to build a visual heat map, that the width of the distribution appears very thin. This occurs regardless of whether we use sodium or silver salt, showing that this not merely an effect of ionisation efficiency, as methacrylate and styrene species ionise more efficiently with different cation species.

Conclusions

The genetic algorithm has been used for automated assignment of copolymer mass spectra, with high accuracy and efficiency. Its utilization on presenting usually complex mass spectra as simple heat maps allows for the qualitative comparison of data, in the case of low molecular weight copolymers. There are still improvements to be made on the implementation of our data processing methodology, such as the way that isotopic distributions are handled means that it is probably ignoring certain overlapping species. To overcome this would either require higher resolution instrumentation, or predicting the amount of intensity within a certain overlapping peak which is to be allocated to each constituent species. Other ways in which the approach discussed here could be altered

would be to allow for the assignment of multiple end groups, as our approach only assigns all copolymer peaks with a given end group. This is simple to overcome in the genetic algorithm methodology; however, it will greatly increase the computational power required to run such a script. The output of having all copolymer peaks assigned, in a simple and automatic manner, allows for a future of more advanced analysis of very complex data sets.

Acknowledgements

James S Town would like to acknowledge the MAS CDT, EPSRC (grant number EP/L015307/1), Syngenta and AstraZeneca for funding and support, as well as Dr Daniel Lester for helpful discussions and the Polymer RTP for use of instrumentation. Christophe J Atkins would like to thank Lubrizol for PhD funding. Ellis Hancox would like to thank the Warwick Monash Alliance for PhD funding. Ataulla Shegiwal would like to thank DSM for PhD sponsorship.

References

1. MacDonald J, Parker M, Greenland B, Hermida-Merino D, Hamley I, Shaver MJPC. Tuning thermal properties and microphase separation in aliphatic polyester ABA copolymers. *Polymer Chemistry.* 2015;6(9):1445-1453.
2. Park S-J, Kang S-G, Fryd M, Saven JG, Park S-J. Highly Tunable Photoluminescent Properties of Amphiphilic Conjugated Block Copolymers. *J Am Chem Soc.* 2010;132(29):9931-9933.
3. Sun L, Liu W, Dong C-M. Bioreducible micelles and hydrogels with tunable properties from multi-armed biodegradable copolymers. *Chem Communs.* 2011;47(40):11282-11284.
4. Katz JS, Doh J, Irvine DJ. Composition-Tunable Properties of Amphiphilic Comb Copolymers Containing Protected Methacrylic Acid Groups for Multicomponent Protein Patterning. *Langmuir.* 2006;22(1):353-359.
5. Shegiwal A, Wemyss AM, Schellekens MAJ, et al. Exploiting catalytic chain transfer polymerization for the synthesis of carboxylated latexes via sulfur-free RAFT. *J Polymer Science.* 2019;57(3):E1-E9.
6. Martin L, Peltier R, Kuroki A, Town JS, Perrier S. Investigating Cell Uptake of Guanidinium-Rich RAFT Polymers: Impact of Comonomer and Monomer Distribution. *Biomacromolecules.* 2018;19(8):3190-3200.
7. Kuroki A, Sangwan P, Qu Y, et al. Sequence Control as a Powerful Tool for Improving the Selectivity of Antimicrobial Polymers. *ACS Applied Materials & Interfaces.* 2017;9(46):40117-40126.
8. Kim H-C, Park S-M, Hinsberg WD. Block Copolymer Based Nanostructures: Materials, Processes, and Applications to Electronics. *Chem Rev.* 2010;110(1):146-177.
9. Blanazs A, Armes SP, Ryan AJ. Self-Assembled Block Copolymer Aggregates: From Micelles to Vesicles and their Biological Applications. *Macromol Rapid Commun.* 2009;30(4-5):267-277.
10. Schmidt K, Grunwald D, Pasch H. Preparation of phenol–urea–formaldehyde copolymer adhesives under heterogeneous catalysis. *Appl Polymer Sci.* 2006;102(3):2946-2952.

11. Lin B, Zhu H, Tieu AK, Hirayama T, Kosasih B, Novareza O. Adsorbed film structure and tribological performance of aqueous copolymer lubricants with phosphate ester additive on Ti coated surface. *Wear.* 2015;332-333:1262-1272.
12. Haddleton DM, Maloney DR, Suddaby A, Clarke KG, Richards SN. Radical-addition-fragmentation and co-polymerization of methyl methacrylate macromonomers from catalytic chain transfer polymerization (CCTP). *Polymer.* 1997;38(25):6207-6217.
13. Kukulj D, Davis TP, Suddaby KG, Haddleton DM, Gilbert RG. Catalytic chain transfer for molecular weight control in the emulsion homo- and copolymerizations of methyl methacrylate and butyl methacrylate. *J Polymer Sci part A: Polymer Chemistry.* 1997;35(5):859-878.
14. Heuts JPA, Kukulj D, Forster DJ, Davis TP. Copolymerization of Styrene and Methyl Methacrylate in the Presence of a Catalytic Chain Transfer Agent. *Macromolecules.* 1998;31(9):2894-2905.
15. Noda T, Grice AJ, Levere ME, Haddleton DM. Continuous process for ATRP: Synthesis of homo and block copolymers. *European Polymer Journal.* 2007;43(6):2321-2330.
16. Darcos V, Haddleton DM. Synthesis of ABABA pentablock copolymers via copper mediated living radical polymerisation. *European Polymer Journal.* 2003;39(5):855-862.
17. Hansen NML, Gerstenberg M, Haddleton DM, Hvilsted S. Synthesis, characterization, and bulk properties of amphiphilic copolymers containing fluorinated methacrylates from sequential copper-mediated radical polymerization. *J Polymer Sci part A: Polymer Chemistry.* 2008;46(24):8097-8111.
18. Batagianni E, Marathianos A, Koraki A, Maroudas A-P, Pitsikalis M. Metallocene-mediated cationic polymerization of vinyl ethers: Kinetics of polymerization and synthesis and characterization of statistical copolymers. *Journal of Macromolecular Science, Part A.* 2016;53(3):140-151.
19. Anderson BC, Andrews GD, Arthur P, et al. Anionic polymerization of methacrylates. Novel functional polymers and copolymers. *Macromolecules.* 1981;14(5):1599-1601.
20. Kuroki A, Martinez-Botella I, Hornung CH, et al. Looped flow RAFT polymerization for multiblock copolymer synthesis. *Polymer Chemistry.* 2017;8(21):3249-3254.
21. Martin L, Gody G, Perrier S. Preparation of complex multiblock copolymers via aqueous RAFT polymerization at room temperature. *Polymer Chemistry.* 2015;6(27):4875-4886.
22. Gody G, Maschmeyer T, Zetterlund PB, Perrier S. Rapid and quantitative one-pot synthesis of sequence-controlled polymers by radical polymerization. *Nature Communications.* 2013;4:2505.
23. Nurumbetov G, Engelis N, Godfrey J, et al. Methacrylic block copolymers by sulfur free RAFT (SF RAFT) free radical emulsion polymerisation. *Polymer Chemistry.* 2017;8(6):1084-1094.
24. Engelis NG, Anastasaki A, Nurumbetov G, et al. Sequence-controlled methacrylic multiblock copolymers via sulfur-free RAFT emulsion polymerization. *Nature Chemistry.* 2016;9:171.
25. Engelis NG, Anastasaki A, Whitfield R, et al. Sequence-Controlled Methacrylic Multiblock Copolymers: Expanding the Scope of Sulfur-Free RAFT. *Macromolecules.* 2018;51(2):336-342.
26. Lotierzo A, Schofield RM, Bon SAF. Toward Sulfur-Free RAFT Polymerization Induced Self-Assembly. *ACS Macro Letters.* 2017;6(12):1438-1443.

27. Vana P, Albertin L, Barner L, Davis TP, Barner-Kowollik C. Reversible addition–fragmentation chain-transfer polymerization: Unambiguous end-group assignment via electrospray ionization mass spectrometry. *J Polymer Sci Part A: Polymer Chemistry.* 2002;40(22):4032-4037.
28. Charles L. MALDI of synthetic polymers with labile end-groups. *Mass Spectrom Rev.* 2014;33(6):523-543.
29. Gruendling T, Hart-Smith G, Davis TP, Stenzel MH, Barner-Kowollik C. Enhanced Ionization in Electrospray Ionization Mass Spectrometry of Labile End-Group-Containing Polystyrenes Using Silver(I) Tetrafluoroborate as Doping Salt. *Macromolecules.* 2008;41(6):1966-1971.
30. Mazarin M, Girod M, Viel S, et al. Role of the Adducted Cation in the Release of Nitroxide End Group of Controlled Polymer in Mass Spectrometry. *Macromolecules.* 2009;42(6):1849-1859.
31. Montaudo G, Garozzo D, Montaudo MS, Puglisi C, Samperi F. Molecular and Structural Characterization of Polydisperse Polymers and Copolymers by Combining MALDI-TOF Mass Spectrometry with GPC Fractionation. *Macromolecules.* 1995;28(24):7983-7989.
32. Gruendling T, Junkers T, Guilhaus M, Barner-Kowollik CJMC, Physics. Mark–Houwink Parameters for the Universal Calibration of Acrylate, Methacrylate and Vinyl Acetate Polymers Determined by Online Size-Exclusion Chromatography—Mass Spectrometry. *Macromolecular Chemistry and Physics.* 2010;211(5):520-528.
33. Martin K, Spickermann J, Räder HJ, Müllen K. Why Does Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry Give Incorrect Results for Broad Polymer Distributions? *Rapid Commun Mass Spectrom.* 1996;10(12):1471-1474.
34. Liu XM, Maziarz EP, Heiler DJ, Grobe GL. Comparative studies of poly(dimethyl siloxanes) using automated GPC-MALDI-TOF MS and on-line GPC-ESI-TOF MS. *J Am Soc Mass Spectrom.* 2003;14(3):195-202.
35. Wickel H, Agarwal S, Greiner A. Homopolymers and Random Copolymers of 5,6-Benzo-2-methylene-1,3-dioxepane and Methyl Methacrylate: Structural Characterization Using 1D and 2D NMR. *Macromolecules.* 2003;36(7):2397-2403.
36. Brar AS, Saini T. Comprehensive 2D NMR analysis: Acrylonitrile/ethyl methacrylate copolymers synthesized by ATRP at ambient temperature. *J Polymer Sci Part A: Polymer Chemistry* 2006;44(9):2955-2971.
37. Aerdts AM, De Haan JW, German AL, Van der Velden GPM. Characterization of intramolecular microstructure of styrene-methyl methacrylate copolymers: new proton NMR assignments supported by 2D-NOESY NMR. *Macromolecules.* 1991;24(7):1473-1479.
38. Weidner SM, Falkenhagen J, Bressler I. Copolymer Composition Determined by LC-MALDI-TOF MS Coupling and “MassChrom2D” Data Analysis. *Macromolecular Chemistry and Physics.* 2012;213(22):2404-2411.
39. Weidner SM, Falkenhagen J. LC-MALDI-TOF Imaging MS: A New Approach in Combining Chromatography and Mass Spectrometry of Copolymers. *Anal Chem.* 2011;83(23):9153-9158.
40. Barqawi H, Schulz M, Olubummo A, Saurland V, Binder WH. 2D-LC/SEC-(MALDI-TOF)-MS Characterization of Symmetric and Nonsymmetric Biocompatible PEOm–PIB–PEOn Block Copolymers. *Macromolecules.* 2013;46(19):7638-7649.

41. Raust J-A, Brüll A, Moire C, Farcet C, Pasch H. Two-dimensional chromatography of complex polymers: 6. Method development for (meth)acrylate-based copolymers. *J Chromatogr A.* 2008;1203(2):207-216.
42. Knecht D, Rittig F, Lange RFM, Pasch H. Multidimensional chromatographic techniques for hydrophilic copolymers: II. Analysis of poly(ethylene glycol)-poly(vinyl acetate) graft copolymers. *J Chromatogr A.* 2006;1130(1):43-53.
43. Hughey CA, Hendrickson CL, Rodgers RP, Marshall AG, Qian K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal Chem.* 2001;73(19):4676-4681.
44. Barrow MP. Petroleomics: study of the old and the new. *Biofuels.* 2010;1(5):651-655.
45. Fouquet T, Sato H. Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units. *Anal Chem.* 2017;89(5):2682-2686.
46. Nakamura S, Cody RB, Sato H, Fouquet T. Graphical Ranking of Divisors to Get the Most out of a Resolution-Enhanced Kendrick Mass Defect Plot. *Anal Chem.* 2019;91(3):2004-2012.
47. Zheng Q, Morimoto M, Sato H, Fouquet T. Resolution-enhanced Kendrick mass defect plots for the data processing of mass spectra from wood and coal hydrothermal extracts. *Fuel.* 2019;235:944-953.
48. Fouquet T, Nakamura S, Sato H. MALDI SpiralTOF high-resolution mass spectrometry and Kendrick mass defect analysis applied to the characterization of poly(ethylene-co-vinyl acetate) copolymers. *Rapid Commun Mass Spectrom.* 2016;30(7):973-981.
49. Nagy T, Kuki Á, Zsuga M, Kéki S. Mass-Remainder Analysis (MARA): a New Data Mining Tool for Copolymer Characterization. *Anal Chem.* 2018;90(6):3892-3897.
50. Fouquet T. Comment on “Mass-Remainder Analysis (MARA): A New Data Mining Tool for Copolymer Characterization” (An Example of Multiple Discovery). *Anal Chem.* 2018;90(14):8716-8718.
51. Willemse RXE, Staal BBP, Donkers EHD, van Herk AM. Copolymer Fingerprints of Polystyrene-block-polyisoprene by MALDI-ToF-MS. *Macromolecules.* 2004;37(15):5717-5723.
52. Wei Z, Vladimir VT. Pattern Recognition and Pathway Analysis with Genetic Algorithms in Mass Spectrometry Based Metabolomics. *Algorithms.* 2009;2(2):638-666.
53. Duan J, Amster IJ. An Automated, High-Throughput Method for Interpreting the Tandem Mass Spectra of Glycosaminoglycans. *J Am Soc Mass Spectrom.* 2018;29(9):1802-1811.
54. Anastasaki A, Nikolaou V, Zhang Q, et al. Copper(II)/Tertiary Amine Synergy in Photoinduced Living Radical Polymerization: Accelerated Synthesis of ω-Functional and α,ω-Heterofunctional Poly(acrylates). *J Am Chem Soc.* 2014;136(3):1141-1149.
55. Liarou E, Anastasaki A, Whitfield R, et al. Ultra-low volume oxygen tolerant photoinduced Cu-RDRP. *Polymer Chemistry.* 2019;10(8):963-971.
56. Jones GR, Whitfield R, Anastasaki A, Haddleton DM. Aqueous Copper(II) Photoinduced Polymerization of Acrylates: Low Copper Concentration and the Importance of Sodium Halide Salts. *J Am Chem Soc.* 2016;138(23):7346-7352.
57. Boyer C, Atme A, Waldron C, et al. Copper(0)-mediated radical polymerisation in a self-generating biphasic system. *Polymer Chemistry.* 2013;4(1):106-112.

58. Burns JA, Houben C, Anastasaki A, Waldron C, Lapkin AA, Haddleton DM. Poly(acrylates) via SET-LRP in a continuous tubular reactor. *Polymer Chemistry.* 2013;4(17):4809-4813.
59. Whitfield R, Anastasaki A, Nikolaou V, et al. Universal Conditions for the Controlled Polymerization of Acrylates, Methacrylates, and Styrene via Cu(0)-RDRP. *J Am Chem Soc.* 2017;139(2):1003-1010.
60. Whitfield R, Anastasaki A, Jones GR, Haddleton DM. Cu(0)-RDRP of styrene: balancing initiator efficiency and dispersity. *Polymer Chemistry.* 2018;9(34):4395-4403.
61. Meyer VE. Copolymerization of styrene and methyl methacrylate. Reactivity ratios from conversion–composition data. *J Polymer Sci Part A: Polymer Chemistry.* 1966;4(11):2819-2830.

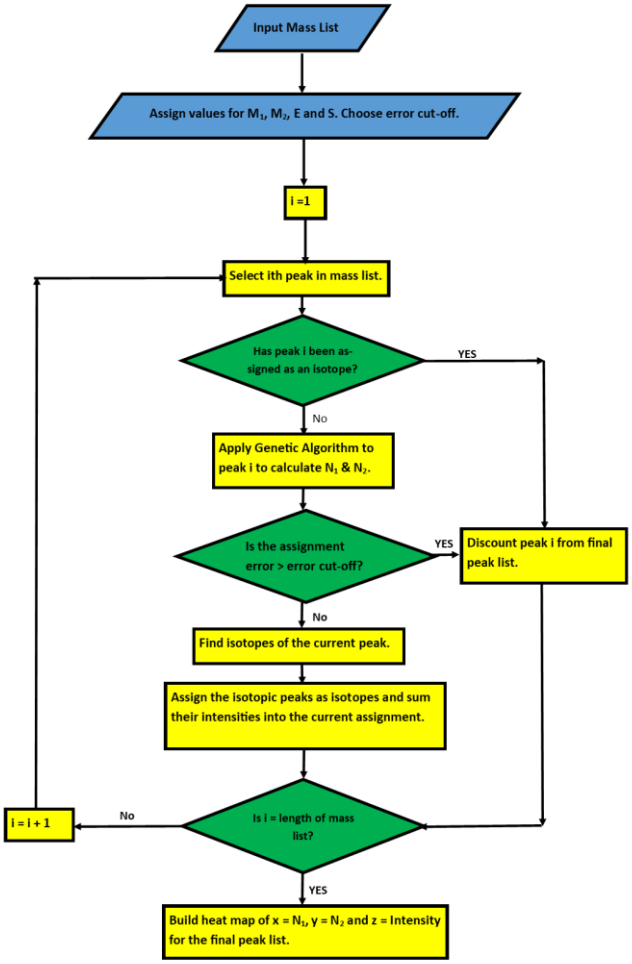

Figure 1: Flow chart of the Matlab script utilized in this paper.

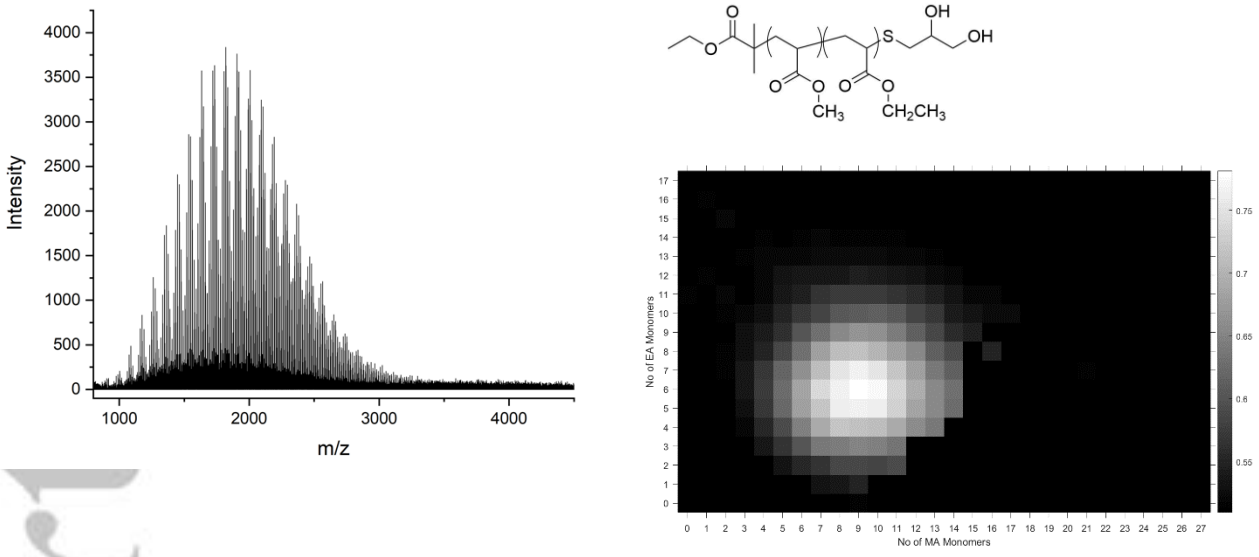

Figure 2: Methyl acrylate - co - ethyl acrylate 50-50 mol% synthesised by photomediated SET-LRP MALDI-ToF results (left), structure of methyl acrylate-co-ethyl acrylate with a substituted thiol end group (top right), heatmap produced from the data (bottom right).

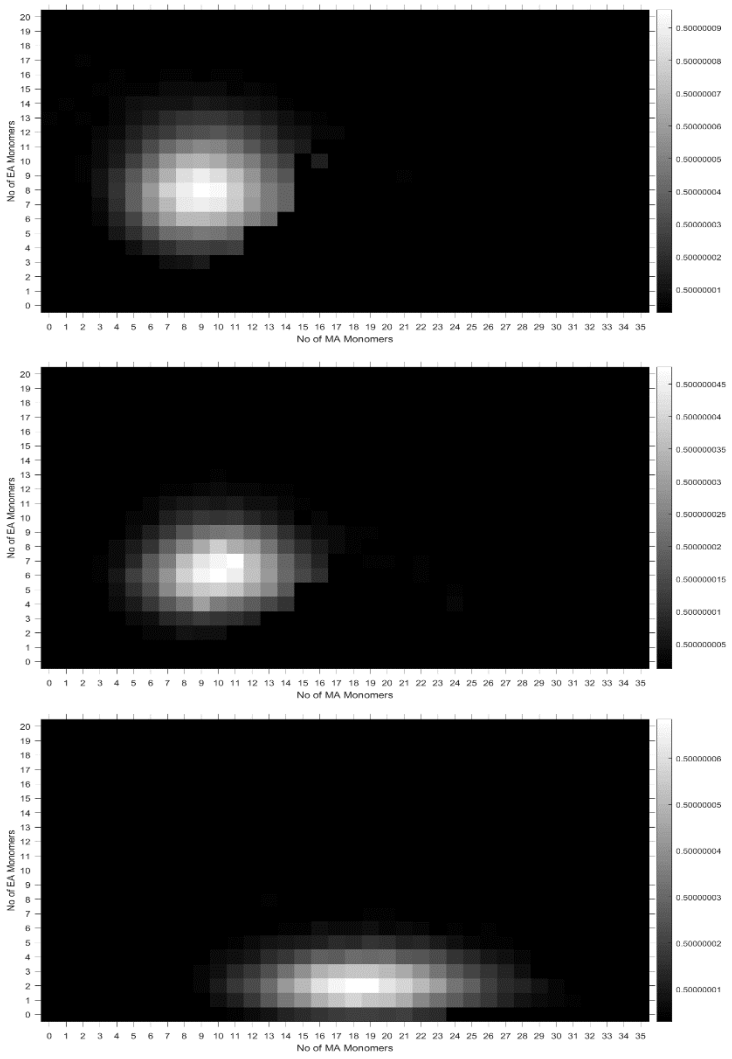

Figure 3: Heat maps visualising the data for methyl-acrylate-co-ethyl acrylate statistical copolymers with ratios of 50-50 (top), 70-30 (middle), 90-10 (bottom).

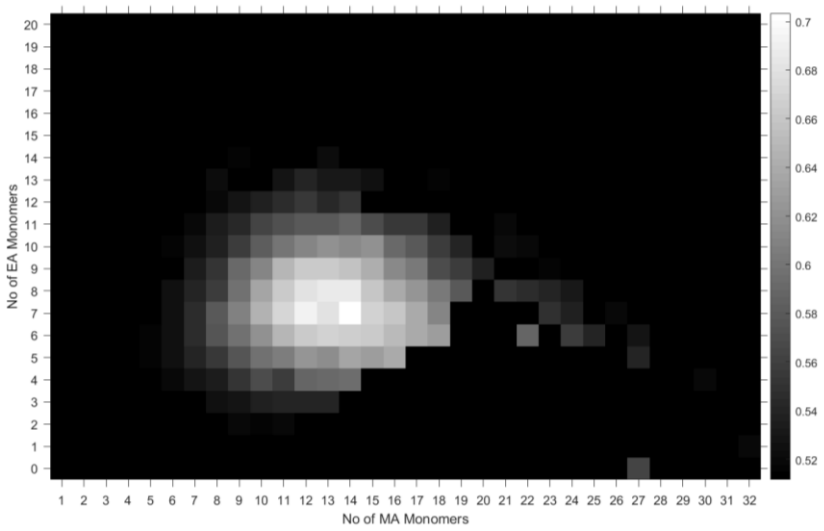

Figure 4: 60/40 Methyl acrylate - co - ethyl acrylate statistical copolymer synthesised by photomediated SET-LRP, showing the issue of miss-assignment of isotopic peaks.

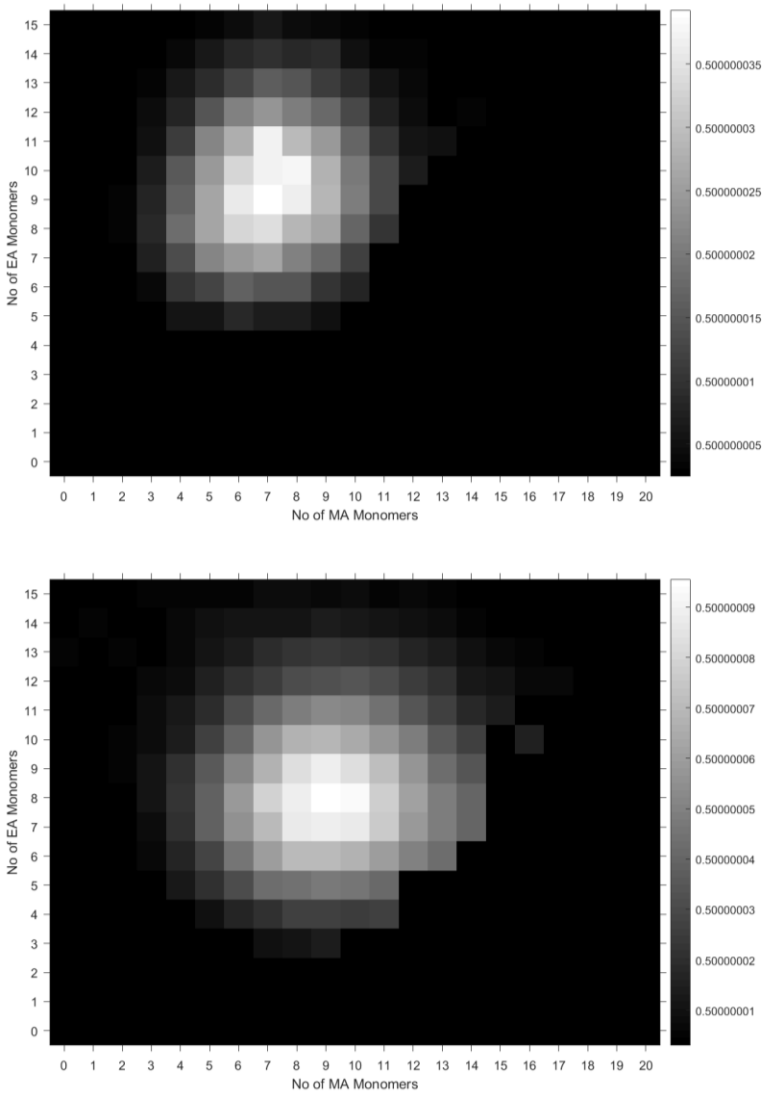

Figure 5: 50-50 Methyl acrylate - co - ethyl acrylate statistical copolymer synthesised by copper(0) wire (top) and photomediated Cu(II) (bottom) SET-LRP.

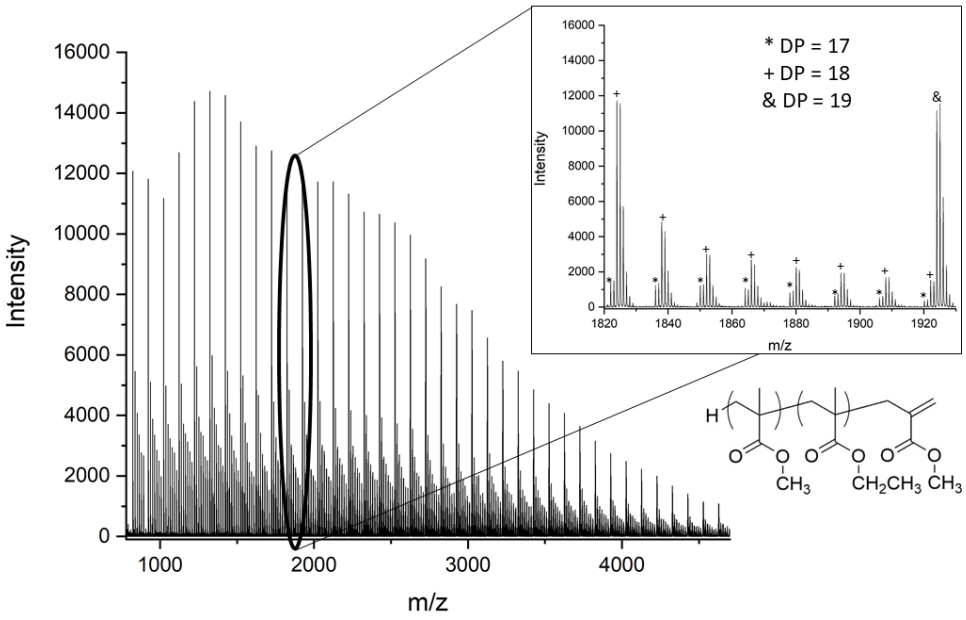

Figure 6: Methyl methacrylate – co – ethyl methacrylate diblock full spectrum, the zoom shows the overlapping of isotopic distributions between different species.

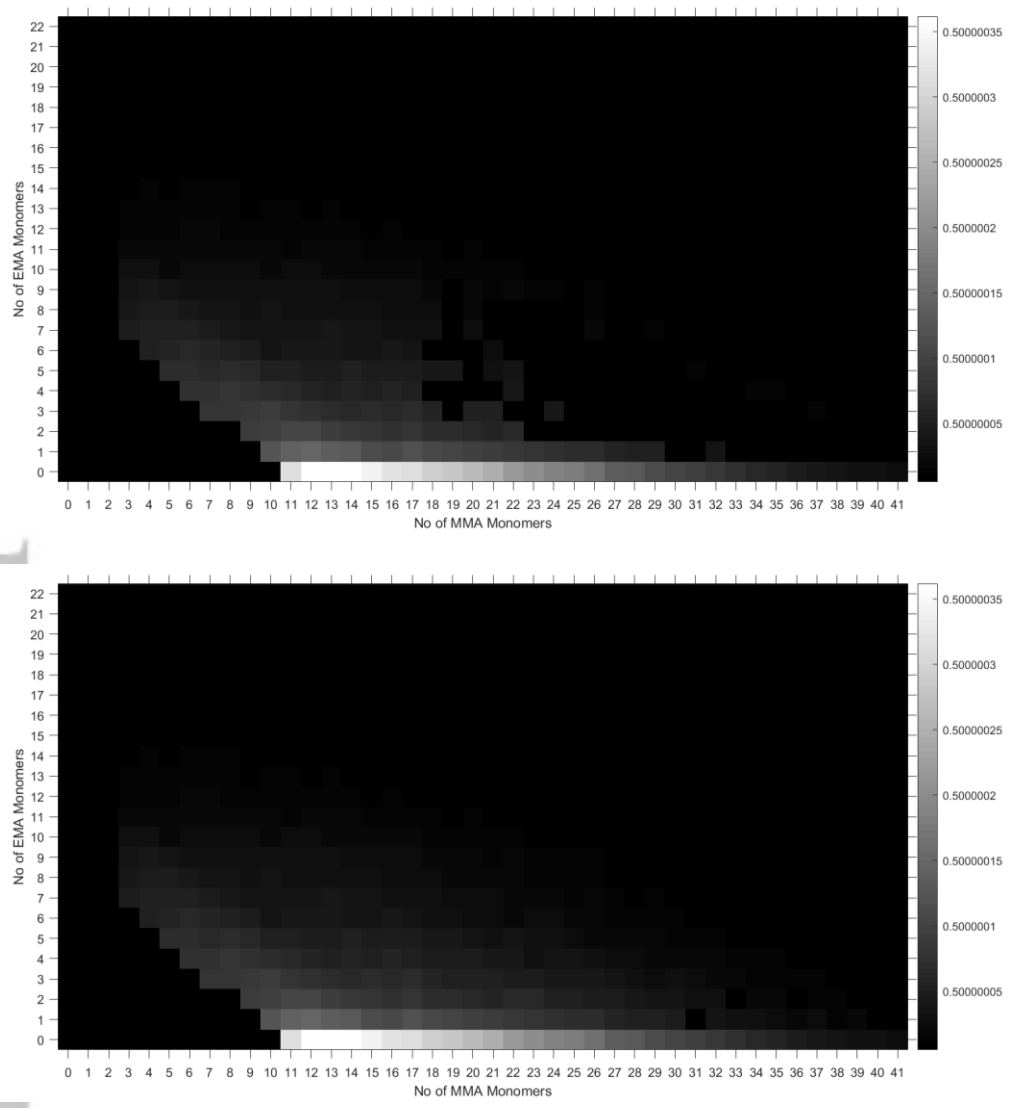

Figure 7: Methyl methacrylate - co - ethyl methacrylate diblock, synthesized by CCTP and SF RAFT. Isotopic intensity issue shown (top), and then resolved (bottom).

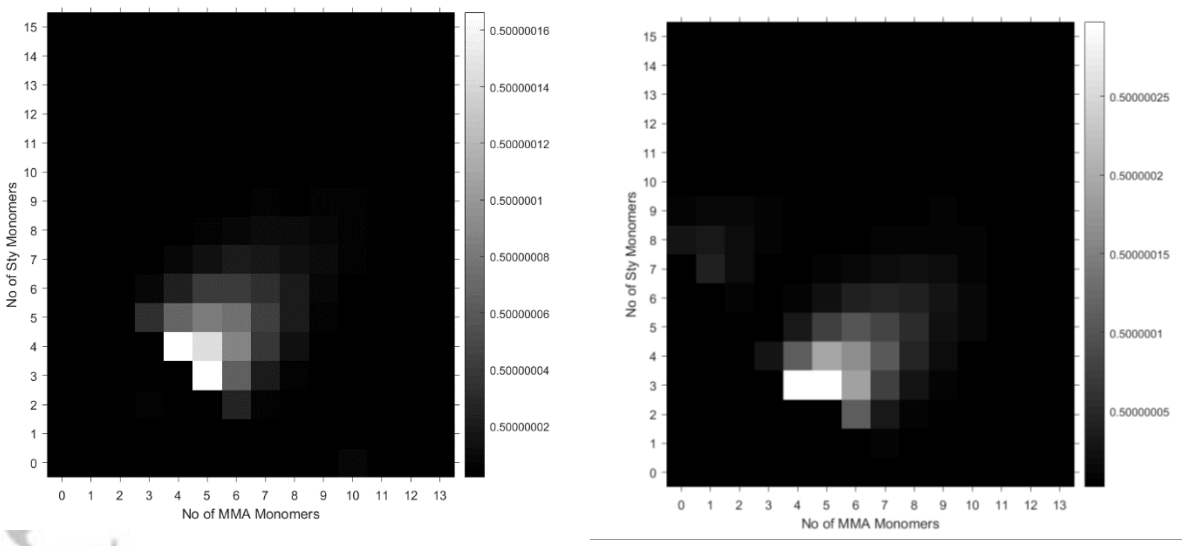

Figure 8: Styrene - co - Methyl Methacrylate statistical copolymer synthesised by bulk free radical, using AgTFA (left) and NaI (right) as a cationising agent.